yang_seonhyeHW6

Seonhye Yang

02/02/2019

```r
library(data.table)
hybrid_reg <- fread("http://users.stat.ufl.edu/~winner/data/hybrid_reg.csv")
hybrid <- data.table("hybrid_reg.csv", header = TRUE)
```

Question 1

```r
attach(hybrid_reg, warn.conflicts = FALSE)
model1 <-lm(msrp~mpgmpge+accelrate, data=hybrid)
summary(model1)
```

```
##
## Call:
## lm(formula = msrp ~ mpgmpge + accelrate, data = hybrid)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38435  -8709  -2836   7755  51093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12309.88    7246.60  -1.699   0.0914 .
## mpgmpge       -131.48      73.85  -1.780   0.0770 .
## accelrate     4740.14     461.21  10.278   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15330 on 150 degrees of freedom
## Multiple R-squared:  0.4945, Adjusted R-squared:  0.4878
## F-statistic: 73.37 on 2 and 150 DF,  p-value: < 2.2e-16
```

Looking at these coefficients, we can get two results. With mpgmpge compared to msrp, we see that the coefficient is -131.48, so for every increase in mile per gallon, the car gets 131.48 dollars cheaper, meaning more fuel efficient cars are cheaper. The opposite is true for acceleration. With a coefficient of 4740.14, for every $4740.14 increase in price, we have a 1 km/hr/sec increase in acceleration, so faster cars are more expensive. Here we have an R^2 value of 0.4878, meaning that only 48.78% of the dependent variable's values can be predicted from the independent variables.

Question 2

```r
rss <- sum(residuals(model1)^2)
rss
```

```
## [1] 35256599813
```

```r
tss <- sum((msrp-mean(msrp))^2)
tss
```

```
## [1] 69747425724
```

```r
mss <- tss - rss
mss
```

```
## [1] 34490825911
```

Question 3

```r
attach(hybrid_reg, warn.conflicts = F)
model2 <- lm(msrp~mpgmpge)
summary(model2)$r.squared - summary(model1)$r.squared
```

```
## [1] -0.3559637
```

Controlling for MPG, and additional 1 km/hr/sec is associated with a lower price of -0.3559637

Question 4

```r
hotel_energy <- fread("http://users.stat.ufl.edu/~winner/data/hotel_energy.csv")
```

```r
attach(hotel_energy, warn.conflicts = FALSE)
fit <- lm(enrgcons~area+age+numrooms+occrate, data = hotel_energy)
fit
```

```
##
## Call:
## lm(formula = enrgcons ~ area + age + numrooms + occrate, data = hotel_energy)
##
## Coefficients:
## (Intercept)         area          age     numrooms      occrate
##  -4692286.2        207.1      15616.5      -7226.2    7366662.2
```

```r
summary(fit)
```

```
##
## Call:
## lm(formula = enrgcons ~ area + age + numrooms + occrate, data = hotel_energy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3627416 -1462717   622058  1052335  2490428
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.692e+06  2.717e+06  -1.727   0.1061
## area         2.071e+02  3.765e+01   5.500 7.82e-05 ***
## age          1.562e+04  1.331e+05   0.117   0.9082
## numrooms    -7.226e+03  6.026e+03  -1.199   0.2504
## occrate      7.367e+06  3.867e+06   1.905   0.0776 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1949000 on 14 degrees of freedom
```

```
## Multiple R-squared:  0.8391, Adjusted R-squared:  0.7931
## F-statistic: 18.25 on 4 and 14 DF,  p-value: 1.92e-05
```

The area as a coefficient of 207.1 which means that as the area increases by one square meters, the energy consumption in kilowatt-hours increases by 207.1kw.

Age has a coefficient of 15616.5 which means that as the age increases by one year, the energy consumption in kilowatt-hours increases by 15616.5kw.

Occupancy Rate in percent has a coefficient of 73666.622 which means that as the occupancy rate increases by one percent, the energy consumption in kilowatt-hours increases by 73666.622kw.

Numrooms has a coefficient of -7226.2 which means that as the number of guestrooms increase by one, the energy consumption in kilowatt-hours decreases by -7226.2kw.

Here we have an R^2 value of 0.79318, meaning that 79.78% of the dependent variable's values can be predicted from the independent variables.

Question 5

```r
controlarea <- lm(enrgcons~age+numrooms+occrate)
summary(controlarea)$r.squared - summary(fit)$r.squared
```

```
## [1] -0.3477274
```

```r
controlage <- lm(enrgcons~area+numrooms+occrate)
summary(controlage)$r.squared - summary(fit)$r.squared
```

```
## [1] -0.0001583283
```

```r
controlnumrooms <- lm(enrgcons~area+age+occrate)
summary(controlnumrooms)$r.squared - summary(fit)$r.squared
```

```
## [1] -0.01652531
```

```r
controloccrate <- lm(enrgcons~area+age+numrooms)
summary(controloccrate)$r.squared - summary(fit)$r.squared
```

```
## [1] -0.04170065
```

The variance for the various parameters is as follows: Area: 0.3477274 Age: 0.00015 NumRooms: 0.0165 Occrate: 0.0417

As we can see, the age predictor accounts for the most variance while controlling for the others.